Abstract of the Disclosure

A method for segmenting video objects in a video sequence that is composed of frames including pixels first assigns a feature vector to each pixel of the video. Next, selected pixels are identified as marker pixels. Pixels adjacent to each marker pixel are assembled into a corresponding a volume of pixels if the distance between the feature vector of the marker pixel and the feature vector of the adjacent pixels is less than a first predetermined threshold. After all pixels have been assembled into volumes, a first score and descriptors are assigned to each volume. At this point, each volume represents a segmented video object. The volumes are then sorted a high-to-low order according to the first scores, and further processed in the high-to-low order. Second scores, dependent on the descriptors of pairs of volumes are determined. The volumes are iteratively combined if the second score passes a second threshold to generate a video object in a resolution video object tree that completes when the combined volume or video object is the entire video.